THE CLAIMS DEFINING THE INVENTION ARE AS FOLLOWS:

1. A method for data tabulation processing of a data file having a plurality of records in a plurality of data fields, comprising:

i) a pre-processing stage in which, for each individual data field, each distinct data value is identified and allocated a numerical identifier unique for that field; and

ii) a tabulation stage in which, for each data record, a cell of a result array is determined based on the numerical identifiers for that record, and the result array cell incremented.

2. A method as claimed in claim 1, wherein the pre-processing stage includes generating from said data file an encoded data file containing the numerical identifiers for the data values in each field, and a mapping file which stores a correspondence between each of the distinct data values in the fields and the corresponding numerical identifiers.

3. A method as claimed in claim 2, wherein a plurality of encoded data files are generated in the pre-processing stage, one for each of the data fields.

4. A method as claimed in claim 2, including generating a tabulation result from said result array and said mapping file.

5. A method as claimed in claim 1, wherein said tabulation stage includes initialising the result array having a number of cells determined by the product of the number of numerical identifiers in the data fields.

6. A method as claimed in claim 1, wherein the tabulation stage includes selecting at least two data fields from the plurality of data fields for tabulation, and generating the result array utilising the numerical identifiers for the selected data fields.

7. A method as claimed in claim 6, wherein for N selected data fields, a cell of the result

array is identified for each data record according to:

$$I = K_1 + D_1K_2 + D_1D_2K_3 + ... + D_1D_2...D_{N-1}K_N$$

where I is the cell identity,

$K_1, K_2, ..., K_N$ are the numerical identifiers for the record in the selected fields, and $D_1, D_2, ..., D_{N-1}$ are numbers of distinct values in the selected fields.

8. A system for data tabulation processing of a data file having a plurality of records in a plurality of data fields, comprising:

i) a coding processor in which, for each individual data field, each distinct data value is identified and allocated a numerical identifier unique for that field; and

ii) a tabulation processor in which, for each data record, a cell of a result array is determined based on the numerical identifiers for that record, and the result array cell incremented.

9. A system as claimed in claim 8, wherein said coding processor generates from said data file an encoded data file containing the numerical identifiers for the data values in each field, and a mapping file which stores a correspondence between each of the distinct data values in the fields and the corresponding numerical identifiers.

10. A system as claimed in claim 8, wherein a plurality of encoded data files are generated by the coding processor, one for each of the data fields.

11. A system as claimed in claim 8, wherein a selection of at least two data fields from the plurality of data fields is made in the tabulation processor, which generates the result array utilising the numerical identifiers for the selected data fields.

12. A system as claimed in claim 11, wherein for N selected data fields, the tabulation processor identifies a cell of the result array for each data record according to:

$$I = K_1 + D_1K_2 + D_1D_2K_3 + ... + D_1D_2...D_{N-1}K_N$$

where I is the cell identity,

K_1, K_2, ..., K_N are the numerical identifiers for the record in the selected fields,

and D_1, D_2, ..., D_{N-1} are numbers of distinct values in the selected fields.